

# Matej Sembera

Sonnenallee 24, Wien, 1220 | semb.mato@gmail.com | +43 664 217 77 41 | github.com/matej897

## ZUSAMMENFASSUNG

---

Ich bin Data Science Student an der Universität Wien mit einem Jahr Erfahrung als Data Science Praktikant. Ich verfüge über fundierte Kenntnisse in den Bereichen Datenanalyse, Business Intelligence, KI und Data Engineering. Ich möchte meine Fähigkeiten in realen Projekten anwenden und gleichzeitig mein Studium fortsetzen. Als EU-Bürger habe ich uneingeschränkte Arbeitserlaubnis. Zudem besitze ich ein Goethe-Zertifikat auf C1-Niveau.

## AUSBILDUNG

---

**Universität Wien**, Master in Data Science Sept 2024 – Laufend

- **Kursinhalte:** Datenethik, Einführung in maschinelles Lernen, Visuelle und explorative Datenanalyse.

**Comenius-Universität Bratislava**, Bachelor in Finanz- und Wirtschaftsmathematik Sept 2021 – Juni 2024

- **Kursinhalte:** Statistik, Volkswirtschaft, Rechnungswesen und Versicherungsmathematik.

## BERUFSERFAHRUNG

---

**Data Science Praktikant**, Allekommen.at – Wien, Österreich Juni 2023 – Juni 2024

- **Erstellung von Dashboards und Datenvisualisierungen (Power BI, Tableau)**
- **Datenanalyse & Reporting (Excel, Python, R, SQL):** Ich habe große Datensätze strukturiert, bereinigt und analysiert. Zudem habe ich Berichte und Dashboards für das Management erstellt und optimiert, Ad-hoc Analysen zur schnellen Beantwortung dringender Fragestellungen durchgeführt.
- **Implementierung von KI (Python, NLP, LLM):** Mitarbeit in einem Team, das ein Modell für Natural Language Processing (OpenAI) integriert hat, sodass das System auch bei Tippfehlern (z. B. ‚Hiarcut‘ statt ‚Haircut‘) relevante Ergebnisse liefert. Dies **senkte die Abbruchrate der Nutzer deutlich**.
- **Entwicklung eines SQL-basierten Suchalgorithmus (Python, SQL, Geodaten):** Entwicklung eines neuen GeoSQL-Suchalgorithmus, der **10-mal schneller** als die vorherige Version ist und hohe Nutzerlast ermöglicht.
- **Erstellung von ETL-Pipelines (Python, Linux):** Automatisierung der Datensammlung mittels Web Scraping, Datenaufbereitung mit Python und Linux-Shell-Kommandos, Speicherung in Datenbank PostgreSQL (**ca. 2.000 Dienstleister hochgeladen**).
- **Optimierung von SQL-Abfragen und Entwicklung in PostgreSQL (SQL):** Analyse und Optimierung von SQL-Abfragen, Erstellung von Funktionen für den Suchalgorithmus, sowie Entfernung von doppelten Daten.
- **Referenz:** Michael Pisnyachevskiy, michael@allekommen.at

## PROJEKTE

---

**Bachelorthesis: Maschinelles Lernen für Aktienindex-Vorhersagen (Note: A)** Link zum GitHub

- In meiner Bachelorarbeit habe ich maschinelles Lernen genutzt, um die Bewegungen des Aktienmarktes für den Folgetag vorherzusagen und darauf basierend gehandelt. Im Testdatensatz (23. Nov 2021 – 24. Jan 2024) übertraf das Modell den S&P 500 um **12,8%**.

**Projekt im Kurs Informationsmanagement & Systemtechnik** Link zum GitHub

- Im Projekt für den Kurs Informationsmanagement haben wir eine Fahrradladen-Website mit JavaScript erstellt, diese mit Docker containerisiert und die Daten von PostgreSQL nach MongoDB mittels .NET migriert.

## ZERTIFIKATE

---

**Associate Data Scientist in Python** Link zum GitHub

- Kenntnisse in Python und Bibliotheken wie Pandas, Matplotlib, Seaborn, NumPy, TensorFlow, Scikit-learn und PyTorch.

**SQL Fundamentals** Link zum GitHub

## IBM Data Engineering Professional Certificate

Laufend

- Ich habe viele ETL-Pipelines mit Python und Linux-Shell-Commands gebaut, mit denen ich Daten aus XML-, JSON-Dateien und Webseiten extrahieren kann. Außerdem habe ich Grundkenntnisse in Werkzeugen wie Airflow, Kafka, Hadoop, Apache Spark und AWS erworben.

## FÄHIGKEITEN

---

**Sprachen:** Deutsch (C1, [Link zum Goethe-Zertifikat C1](#)), Englisch (C1, [Link zum IELTS](#)), Slowakisch (C2)

**Soft Skills:** Analytisches Denken, hohe Eigenverantwortung, strukturierte Herangehensweise an komplexe Probleme.

**Programmiersprachen:** Python, R, SQL.

**Werkzeuge:** Excel, Power BI, Tableau, GitHub, Docker, Power Query, Airflow, Kafka, Hadoop, Apache Spark.

**Datenbanken:** PostgreSQL, MongoDB, MySQL.

**Datenformate:** XML, JSON.

**Cloud-Plattformen:** AWS.

**MS Office:** Gute Kenntnisse.